

割引基準マルコフ決定理論

著者	船木 洋一
号	5
発行年	1980
URL	http://hdl.handle.net/10097/14771

ふな き よう いち
船 木 洋 一

学位の種類 経済学博士

学位記番号 経博第5号

学位授与年月日 昭和56年1月16日

学位授与の要件 学位規則第5条第1項該当

研究科・専攻 東北大学大学院経済学研究科
(博士課程) 経済学専攻

学位論文題目 割引基準マルコフ決定理論

論文審査委員 (主査)
教授 竹内 清 助教授 高橋 幸雄
助教授 細谷 雄三

論文内容の要旨

今までのマルコフ決定理論には、応用上のいくつかの問題点がある。本論文は、主としてそれを克服するための新しい理論を扱っている。第1章はマルコフ決定理論の問題点と各章の関係を述べている。

マルコフ決定理論は大変幅の広いモデルであり、その応用分野は、生産管理、在庫管理、取り替え問題、自動制御など多岐にわたっている。マルコフ決定理論のモデルは大別して割引率がある場合とない場合があるが、本論文では割引率が在る目的関数をもった最適化問題を扱っている。

割引のあるマルコフ決定理論には通常、次のような基本的な仮定がある。

1. 状態推移のマルコフ性
2. 各状態での行動の独立性
3. 推移確率や割引率の定常性

1の性質からマルコフ決定理論と名づけられている。上記の仮定が成り立たないときには直接このモデルを使うことはできない。本論文はそのようなときの変形モデルを扱っている。

第Ⅱ章は、この割引のあるマルコフ決定理論に関して、簡単なサーベイを行なった。

第Ⅲ章では、上記の基本的な仮定と直接、関係はないが、時間遅れのあるモデルを扱っている。二つの場合を扱っているが一つは行動に時間遅れがある場合で、もう一つはシステム

の状態に関する情報に時間遅れがある場合である。近年、情報の価値ということが言われているが、これらのモデルを使えば、行動や情報の遅れと費用を結びつけた場合を扱うことができる。

経費を節減すると、行動や、情報の伝達時間に遅れがでる。このとき、経費を節減した方がよいだろうか。逆に、お金をかけてそのような遅れをなくすることができるとき、お金をかけた方がよいだろうか。このような問題に対する判断のものを与えることができる。

ベルマンの最適性の原理より、関数方程式を導き、この方程式の構造から状態空間をつくり直し、通常のマルコフ決定理論へと帰着する。その結果、行動に遅れのある場合と、情報に遅れのある場合が同じモデルになる。

第Ⅳ章は上記の仮定 3 と関連する。通常割引のある目的関数では、すべての期間にわたって、割引率は一定と仮定される。ところが、現実にはそうではなく、変動する方が一般的である。そこでこの第Ⅳ章では期待割引総報酬基準で、割引率が変動するモデルを考える。前半は割引率がそれ自身、マルコフ的に変動する場合を考える。それは現在の割引率が高いときに、ある割合で次期の割引率は低くなり、残りの割合でさらに高くなるというような関係がある場合に当てはまるモデルである。この場合は状態空間を拡張することにより通常のモデルに帰着する。

第Ⅳ章の後半部分は、全く割引率がでたらめに変動する場合を考える。この場合、最適政策を求めることは一般的に不可能である。しかしある一定の範囲内で変動するのであれば、割引率一定の場合と同様の最適政策が存在するということを証明する。その結果は、割引率の変動がある一定範囲内と考えられるとき、割引率一定と仮定して最適政策を求め、それを用いる理論的根拠となるものである。

第Ⅴ章は上記 2 の仮定と関連する。マルコフ決定理論では、個々の状態における意志決定が、他の状態の意志決定に影響を与えない、ということが仮定されている。ところがその条件にあてはまらない一つの興味あるモデルがある。それを第Ⅴ章で扱う。そこでは、たとえば、保険に加入した方がよいか、しない方がよいかの簡単な比較の基準が与えられる。

ところで、これまでのマルコフ決定理論では、目的関数として期待割引総報酬基準と、期待時間平均基準が主としてとりあげられてきているが、直面する問題によっては、それらの目的関数では不適当な場合がある。たとえば、ある時間 N まで C だけの割引総報酬を得なければ会社が倒産してしまうような場合を考えてみよう。期間が $0, 1, 2, \dots$ と番号づけられているとして、每期、每期、確実に $r \left(< \frac{C}{N+1} \right)$ を得ることのできる政策があるとする。今、簡単のために割引率 $\beta = 1$ であるとしておこう。一方不確実であってさらに N 期までの期待割引総報酬が $(N+1)r$ よりも小さいけれど、 N 期までの割引総報酬が C を越える可能性のある政策があるとする。期待割引総報酬を最大にする政策を選択する、という基準で

あると前者を採用してしまう。このような倒産するか、しないかの場合には、いくらかでも倒産をまぬがれる可能性のある政策を採用するのではないだろうか。

そこで第IV章では、このような問題に対して、“割引総報酬が、ある一定の水準を超える確率を最大にする政策を選ぶ”という基準を提唱する。この基準と先の二つの基準の三つを考えておけば、マルコフ決定理論の守備範囲はかなり広がるように思われる。

このモデルは状態空間が連続無限集合のモデルとなる。それで最適解の存在証明や、その最適解を求めるアルゴリズムがむずかしくなる。とくに、最適な政策は、一般的には求めることができない。この章の前半では、この基準に対して、定常な最適政策が存在することを証明し、後半では、最適な定常政策を求めるアルゴリズムを述べた。

最後に、マルコフ決定理論のモデルに典型的にあてはまるものとして、在庫管理のモデルがある。この在庫管理のモデルは古くから扱われてきているものであるが、とくに推移確率や直接報酬（実をいうと損失であるが）にはきわだった特徴があり、そこから、このモデルに対しては最適政策にも特徴があらわれてくる。とくに動的な問題の最適政策だけがもつ、典型的な特徴を見ることができる。その大きな一つは計画期間に対する最適政策の依存である。そこで第VII章では（ s 、 S ）政策と呼ばれる最適政策が存在する古典的なモデルを考え、このパラメーター（ s 、 S ）のとくに S の計画期間への依存関係を調べた。

この S は最大の在庫保有量とも考えられるものであるが、発注時に固定費がかからないときには、計画期間の増大とともに増大してくることが知られている。そえすと直観的に発注時に固定費がかかれば、ますます、増大する傾向にあるのではないかと考えられるが、在庫費用その他のかねあい、単純にそうはならない。それで、この S が計画期間の増大とともに、増大するモデルを考えて、保有在庫量と、計画期間や、他のそれに影響を与えるパラメーターとの関係を明らかにしようとしたものである。

そうして得られた一つの結論は、不確実性が、問題の重要な要因になっており、計画期間が長くなることによって、一層、不確実性が増すのであれば、 S も単調に増大する傾向にあるということである。

論文審査結果の要旨

I 時刻とともに確率的に変動する環境下における意思決定のモデルとして最も簡単で標準的なものの一つ、マルコフ決定過程は、R. A. Howard の先駆的な研究以来、20余年の間に多くの人々によって研究が進められ、その基本的性質はかなり知られるようになった。しかし、その応用となると、潜在的可能性は認められながらも、モデルがいろいろな面

理想化されているため、十分に行われているとは言い難い。本論文は、著者がマルコフ決定モデルの応用範囲を広げようと、現実と従来のモデルの間のギャップを埋めるための努力をした結果をまとめたものである。本論文において、著者は、経済学・経営学の分野における応用を念頭に置きつつ、マルコフ決定過程的ではあっても従来のマルコフ決定理論そのままでは扱い得なかったいくつかのタイプの問題を考察し、新しいモデルを提案してその性質を解明した。

II 第1章は本論文で扱った問題およびモデルと各章の簡単な紹介に充てられている。

第2章はマルコフ決定理論、とくに割引率をもったマルコフ決定理論、に関する従来の主要な研究のサーベイである。R. Bellman によって導入された動的計画法の考えが R. A. Howard によってマルコフ過程と結びつけられ、その後 C. Deman, F. D'Epenoux D Blackwell, A. F. Veinott, Jr らによって発展させられてきた道筋が簡潔にまとめられている。また割引率と定常政策との関係についての研究や、各種の変形モデルも紹介されている。

第3章では、時間遅れのあるマルコフ決定問題を解明する目的で、行動に1期の時間遅れのあるモデル(3.1節)、行動にn期の時間遅れのあるモデル(3.2節)、状態についての情報に1期遅れのあるモデル(3.3節)、状態についての情報にn期の遅れのあるモデル(3.4節)という4種のモデルを定式化し、そのいずれもが、状態空間をはじめ、行動空間、直接報酬、推移確率などを適当に定義し直すことによって通常のマルコフ決定モデルに帰着できることを示している。これらのモデルは、決定の影響が現れるのに時間がかかる場合および真の状態が何であるかを観測するのに時間がかかる場合に有効である。

従来のマルコフ決定理論では、ごく一部の例外を除いて、割引率は時刻に依存せず一定であると仮定していた。しかし利子率を用いて割引率に何らかの意味付けをした場合、この仮定は実際の経済現象にあてはまらない。第4章ではこの問題を取りあげ、割引率が時刻とともに変動したときの最適政策の性質、とくに定常最適政策が存在するかどうか、について解析を行っている。まず割引率自体がシステムの状態などとは独立にマルコフ的変動をする場合を考え、システムの状態と割引率の状態を組にしたものを新たな状態とすることにより従来のマルコフ決定モデルに帰着されることを示した(4.1節)。つぎに割引率がある区間で変動するならば最適な定常政策が存在することを示し(4.2節)、さらに割引率が単調に変化する場合に最適な定常政策が存在するための十分条件を与えている(4.3節)。

保険をかけるか否かの決定プロセスは数学的にはある種の賭のモデルとして定式化するのが自然である。第5章ではこのモデルがある特殊なタイプのマルコフ決定過程モデルに

定式化できることを示し、その最適解を求めるのに都合のよい条件を示している。

第6章は著者の創意が最も良く現われた部分で、確率基準を用いる新しいタイプのマルコフ決定過程モデルを提唱し、その性質および最適政策の近似解の求め方について明らかにしている。このモデルは、従来の“報酬の割引現在価値の期待値”を最大にする替りに、“報酬の割引現在価値がある与えられた水準を越える確率”を最大とする政策を求めるもので、この場合、目的関数は各期の報酬に対して非線形となり数学的に相当難しくなる。著者は近似解を巧みに利用することによりこの難しいモデルの解析に成功している。

6.1節では、一定の利益があがらなければ破産してしまう場合など、この確率基準が適当である現象をあげ、他の期待値基準、モーメント基準との比較を行っている。6.2節では、無限計画期間の確率基準マルコフ決定モデルを定式化し、ごく緩い条件の下で最適な定常政策が存在することを、一連の定理と補題を証明することにより示している。6.3節では、報酬の割引現在価値がとり得る値をいくつかの区間に分け、各区間をその中の1点で代表させることにより確率基準マルコフ決定モデルを近似する期待値基準マルコフ決定モデルが作れることを示し、6.4節で、その区間の幅を小さくすることによって近似度を望むだけ向上させられることを示している。これにより、任意の精度で最適な定常政策を近似する政策を数値的に計算して求めることが可能となった。6.5節では、計画期間が有限の確率基準マルコフ決定モデルについてその性質を調べ、無限計画期間の場合との関連について議論している。

第7章では、マルコフ決定理論の代表的な応用例の一つである在庫管理モデルを取りあげ、その計画期間の長さ N と、 (s, S) 型政策の発注量に関する特性量 S の値との関係を調べている。発注時に固定費がかからない場合には N が増大するにつれて S の値も単調に増加することが知られているが、ここではこの結果を固定費がかかる場合にまで拡張することを阻っている。7.1節で多段階在庫管理モデルを説明するとともに、従来の主要な研究の結果をまとめ、7.2節である条件が満たされれば S は単調性をもつことを示している。この条件はそれほど一般的なものでなく、この結果は、むしろ S の単調性が成り立たない場合が多いことを示唆しているようにも思える。

Ⅲ 本論文において著者は一貫してマルコフ決定過程を研究し、経済学・経営学の分野における応用を意図して従来のモデルと実際の現象の間のギャップを埋める新しいモデルを各種提案するとともに、それらの性質を明らかにしている。モデルの仮定にはまだ特殊すぎるものもあり、また解析の点でも十分でない部分が散見されるが、具体的な現象を念頭に置いて作られたこれらのモデルは学界のみならずこれらのモデルを利用する諸分野に貢献をなし得るものと考えられる。

理論的にみると、割引率が変動するモデルと確率基準を用いるモデルが興味深い。これらの問題が、動的計画法の枠組の中でその数学的解明ができることを示したことは、今後のマルコフ決定理論の拡張に道を拓いたものであり、その意義は大きい。とくに確率基準を用いるモデルは目的関数を従来の期待値から確率に置き換えたもので、適用可能な現象の範囲も広く、また多くの変形モデルも導出可能である。

以上により、本論文は経済学博士論文として合格と認定する。